THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 263, No. 30, Issue of October 25, pp. 15335-15341, 1988
Printed in U.S.A.

Isolation and Complete Sequence of a Functional Human Glyceraldehyde-3-phosphate Dehydrogenase Gene*

(Received for publication, March 14, 1988)

Louis Ercolani‡, Brian Florence, Maria Denaro, and Maria Alexander

From the Department of Medicine and Howard Hughes Medical Institute Laboratories, Harvard Medical School and the Medical Services and Diabetes Unit, Massachusetts General Hospital, Boston, Massachusetts 02114

Glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.2.12) (GAPDH) mRNA levels, protein, and enzymatic activity increase in 3T3-F442A adipocytes after exposure to physiological concentrations of insulin (Alexander, M., Curtis, G., Avruch, J., and Goodman, H. (1985) *J. Biol. Chem.* 260, 11978–11985). In order to understand the mechanism of this regulation, we have isolated and sequenced 5.4 kilobase pairs of a 12-kilobase pair human genomic clone encoding a functional GAPDH gene. The gene consists of 9 exons and 8 introns with eukaryotic signals necessary for the transcription and translation of GAPDH mRNA. The exon sequence confirms previously published cDNA sequences for human GAPDH in muscle, liver, and erythrocytes. The organization of the human and the unique chicken GAPDH genes is strikingly similar. Although chicken exons VIII–XI have been fused into human exon 8, introns which separate exons encoding the NAD binding, catalytic, and helical domains of the GAPDH protein have been retained. Stable transfection of rodent cells with the intact human GAPDH gene resulted in the expression of a correctly initiated human GAPDH mRNA and an enzymatically active human GAPDH polypeptide. Thus, the gene contains a functional promoter and intact coding sequences. Although many processed GAPDH pseudogenes and GAPDH-like sequences are present in the human genome, Southern blot analysis of human genomic DNA using a probe derived from the 3′-untranslated region of the GAPDH gene detected only two genes, a 10-copy processed pseudogene and a single copy of the isolated gene. In contrast, a probe derived from an intron segment of the isolated gene detected only a single copy of the GAPDH gene. Collectively, these findings strongly suggest that the human genome encodes a single functional GAPDH gene.

The glycolytic enzyme glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.2.12) (GAPDH)[1] catalyzes the conversion of glyceraldehyde 3-phosphate to 1,3-diphosphoglycerate (1–3). GAPDH is a tetramer composed of identical subunits of 37,000 molecular weight (2, 3). The activity of the enzyme is regulated by several glycolytic metabolites and by the association of the enzyme with cell membranes (4, 5). Our interest in GAPDH stems from recent observations that insulin increases GAPDH gene transcription in 3T3-F442A adipocytes[2] which results in a proportionate increase in GAPDH mRNA levels and enzymatic activity in these cells (6). This effect and other effects of insulin on diverse metabolic processes is initiated through the interaction of insulin with a ligand-specific cell surface receptor with intrinsic tyrosine kinase activity (7). The events subsequent to this interaction that result in alterations in gene expression have yet to be elucidated. As current concepts suggest that alterations in gene transcription may result from a direct interaction of *trans*-acting factors with *cis*-acting sequences in hormone-regulated genes, utilizing the sequences responsible for insulin-mediated regulation of this gene should allow us to isolate and identify *trans*-acting factors that mediate the effect of insulin on GAPDH gene transcription. Thus, we sought to isolate a functional GAPDH gene.

Although somatic hybrid studies have localized a functional GAPDH gene to chromosome 12 and no isoenzymatic forms of the gene with glycolytic activity have been described (8), recent reports have demonstrated that GADPH belongs to a large multiple gene family. This gene family contains 150 or more GAPDH-like sequences, many of which are processed pseudogenes (9, 10). Thus far, attempts to clone the human GAPDH gene have resulted in the isolation of two processed pseudogenes for GAPDH localized to the X chromosome (11, 12). In this paper, the primary structure of a 5.37-kilobase segment of genomic DNA containing the functional human GAPDH gene is presented.

EXPERIMENTAL PROCEDURES AND RESULTS[3]

DISCUSSION

This report describes the isolation and characterization of a human genomic DNA segment which contained a functional GAPDH gene. The isolated gene satisfied our screening criteria insofar as it is the only DNA segment in the human genome that contained both introns and sequences similar to the extreme 3′-untranslated region of the expressed human GAPDH mRNA (see Fig. 3). As seen in Fig. 8, *lanes A* and *B*, Southern blot analysis of human genomic DNA performed with a probe derived from intron B of the isolated gene detected one copy of the GAPDH gene. In contrast, a probe

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J04038.

‡ Supported by National Institutes of Health Grant AM7028.

[1] The abbreviations used are: GAPDH, glyceraldehyde-3-phosphate dehydrogenase; kb, kilobases; pghGAPDH, 12-kb human genomic segment encoding GAPDH subcloned into pUC12; SDS, sodium dodecyl sulfate.

[2] M. Alexander and H. M. Goodman, unpublished results.

[3] Portions of this paper (including "Experimental Procedures," "Results," Figs. 1–7, and Table I) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are included in the microfilm edition of the Journal that is available from Waverly Press.

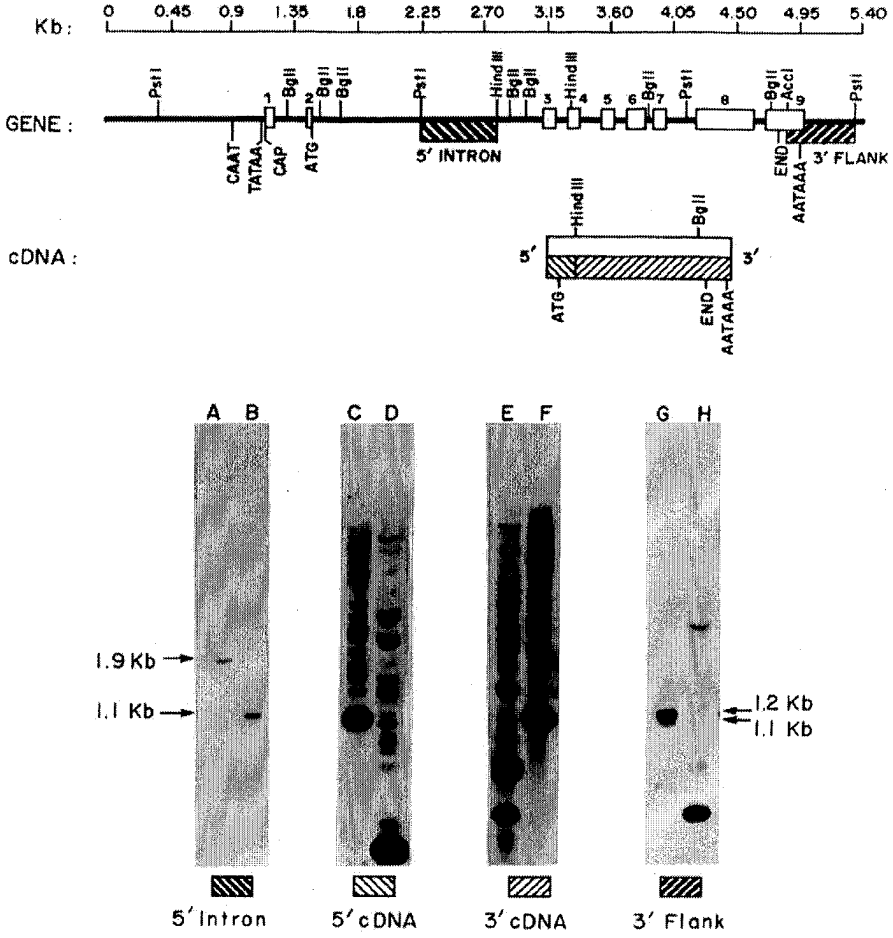

FIG. 8. **Southern blot analysis of the human GAPDH gene.** The structure of the human GAPDH gene is shown with restriction enzyme sites used for Southern analysis; *open boxes* correspond to the exon segments in the human GAPDH gene or the human GAPDH cDNA. *Closed bars* indicate intron segments in the GAPDH gene. DNA isolated from human placenta was digested with restriction enzymes and subjected to Southern blot analysis as described under "Experimental Procedures." ▩ corresponds to the intron fragment used to hybridize human placental DNA digested with *Pst*I in *lane A* and *Hin*dIII-*Bgl*I in *lane B*. ▩ corresponds to the 5′-cDNA GAPDH fragment used to hybridize human placental DNA digested with *Pst*I in *lane C* and *Hin*dIII-*Bgl*I in *lane D*. ▨ corresponds to the 3′-cDNA GAPDH fragment used to hybridize placental DNA digested with *Hin*dIII-*Bgl*I in *lane E* and *Pst*I in *lane F*. ▨ corresponds to the 3′-flanking fragment of the human GAPDH gene used to hybridize placental DNA digested with *Pst*I in *lane G* and *Hin*dIII-*Bgl*I in *lane H*.

derived from the 3′-untranslated region of this gene detected one copy of the GAPDH gene and 10 copies of a processed GAPDH pseudogene (*lanes G* and *H*). We believe this pseudogene is the same DNA segment cloned by Hanauer and Mandel (12) because the exon portion of our 3′-flanking probe was sufficient in length (84 bases) and in sequence similarity (96%) to detect this pseudogene. This finding indicated that there are only two genes in the human genome with sequences similar to the 3′-untranslated region of the GAPDH gene expressed in human tissues, the gene we have isolated and a processed pseudogene. For this reason, we are certain that we have cloned the unique GAPDH gene.

Furthermore, the size of the active transcriptional unit was consistent with that predicted by Dani *et al.* (41) for the human GAPDH gene expressed in HeLa cells. An active promoter was located within 112 base pairs 5′ to the start site of transcription of this gene. The mRNA encoded by this human DNA segment was correctly initiated (see Fig. 4). In studies published elsewhere, we have shown that this mRNA directed the synthesis of a full length polypeptide (38). Here we have demonstrated that after stable integration and expression of the gene in H35 hepatoma cells, an enzymatically active human GAPDH polypeptide was produced (see Fig. 5). Thus, we have isolated a functional GAPDH gene.

Nevertheless, we and others (9–12) have detected a very large number of other GAPDH-related sequences in the human genome with probes derived from the GAPDH cDNA coding region (see Fig. 8). Several groups have speculated about the functional significance of these genes. Acari *et al.* (36) have suggested that a fetal form of the GAPDH gene exists; however, their report was not confirmed by Fort *et al.* (34). The origin of the other GAPDH-related sequences remains unclear. It is likely that some may represent segments of other genes, such as kinases and dehydrogenases, which have acquired NAD binding or catalytic domains similar to those in the GAPDH gene through exon shuffling (42, 43). Alternatively, others may represent genes with structures similar to GAPDH, that perform different functions.

By analyzing glycolytic enzyme activity in somatic hybrids, Bruns *et al.* (8) localized the GAPDH gene to chromosome 12. However, in certain tissues, GAPDH is not a cytosolic protein and appears to mediate functions other than glycolysis. For instance, Tsai *et al.* (5) demonstrated that GAPDH bound to Band 3, the anion transport protein in erythrocyte cell membranes, is enzymatically inactive. Furthermore, GAPDH may have novel functions which are necessary for the formation of specialized membrane structures in brain (44) and skeletal muscle transverse tubules (45). For instance, Kawamoto and Caswell (46) have suggested that in rabbit skeletal muscle, GAPDH may function as a protein kinase. While glycolytic isoenzymes of GAPDH have not been found (8), the presence of a GAPDH protein which performs diverse functions in certain tissues raised the possibility that other GAPDH genes might exist in the human genome. However,

the recent cloning of p37, an intracellular membrane-associated protein in human erythrocytes which is expressed on the extracellular surface of K562 erythroleukemia cells, unexpectedly revealed this protein to be identical in sequence with the cytosolic form of GAPDH (37). Furthermore, comparison of the human GAPDH gene exon sequence to the combined sequences of GAPDH cDNAs isolated from human adult liver, muscle, and erythrocyte suggested complete identity (see Fig. 3). Thus, it appears the GAPDH gene we have cloned encodes the GAPDH protein that mediates both glycolytic and nonglycolytic functions in all human tissues.

GAPDH is present in both prokaryotes and eukaryotes and is highly conserved. Such conservation makes the GAPDH gene a useful model for the study of evolutionay function of intervening sequences. Based on their studies of the chicken GAPDH gene sequence, Stone *et al.* (13) have proposed that the ontogeny for this highly conserved protein is the result of duplication events involving ancestral introns as well as exons. The conservation of introns separating domains predicted by these authors to occur in GAPDH genes of higher eukaryotes has been confirmed. The location of three intervening sequences which separate exons encoding four protein domains of the GAPDH protein detected by x-ray crystallography has been retained in the human GAPDH gene (47, 48). Specifically, intron IV in the chicken gene (intron D in the human gene) divides the two mononucleotide-binding domains; intron VI in the chicken (intron F in the human gene) separates the dinucleotide-binding domain (amino acids 1–149) from the catalytic domain (amino acids 149–312); and intron XI in the chicken gene (intron H in the human gene) separates the helical domain (amino acids 313–334).

Three introns in the human gene corresponding to chicken introns VIII-X have been lost in tandem. Random intron deletions have been described in higher eukaryotic genes (49–51). However, the finding of a 10-copy processed GAPDH pseudogene suggested a possible recombinant event between a retroviral cDNA and a GAPDH gene similar in structure to the chicken GAPDH gene. Based on a predicted mutation rate of 0.7% per million years in the nonfunctional GAPDH pseudogene sequence (51), this event would have occurred within the past 5.7 million years. Thus, our hypothesis would predict an intron/exon organization similar to that of the chicken GAPDH gene in primates such as baboons which diverged from pre-human primates before the GAPDH pseudogene was produced, whereas random deletions of these introns could have occurred at any time after the avian radiation of 270 million years ago.

The 5′- and 3′-flanking regions of this gene contain many repetitive elements. Eleven direct repeats were found in the 5′-flanking region ranging from 7 to 19 bases long. Data base searches revealed several sequences that may contribute to the regulation of the human GAPDH gene. Inverted repetitive elements CCGCCC-GGCGGG associated with upstream enhancer elements for the human and mouse metallothionein genes (52–54) were present in position −972 to −950. A core enhancer sequence GTGGAAG described by Weiher *et al.* (55) was present on the noncoding strand of intron G. "CAAT" box sequences are usually located within 100 bases of the TATAAA sequence (56). A "CAAT" box sequence GGGCCAATCT, present in virtually all globin genes, was located 203 bases upstream of the TATAAA sequence in the human GAPDH gene. However, analysis of constructs with deletions of the 5′-flanking sequences had no more than a 50% decrease in promoter activity when the CAAT sequence was removed.

In studies published elsewhere, we have shown that the human GAPDH gene is regulated by insulin when transfected and expressed in differentiated F442A adipocytes but not preadipocytes (38). The differentiation of preadipocytes to adipocytes is stimulated by agents which demethylate chromatin (57, 58). It is therefore of some interest that the GAPDH promoter was found in an area (−630 to −1 nucleotides) that is 65% G-C-rich, with 41 CpG dinucleotides and 59 GpC dinucleotides. This area is similar to the well described *HpaII* tiny fragment islands that are characteristic of mammalian housekeeping genes (59). It is possible that this *HpaII* tiny fragment island with 41 potential methylation sites may regulate GAPDH gene expression during the differentiation of preadipocytes. The availability of a probe which detects a unique GAPDH gene rather than the numerous pseudogenes present in the mammalian genome will make it possible to investigate whether methylation influences GAPDH gene expression during differentiation.

Having isolated the human GAPDH gene, it will be feasible to delineate the sequences and characterize the factors responsible for the transcriptional activation of this gene by insulin. These studies are currently in progress in our laboratory. We believe the isolation of the human GAPDH gene will afford many opportunities for understanding both gene expression in general and the final effector pathways for insulin action in particular.

Acknowledgments—We are indebted to H. M. Goodman in whose laboratory the isolation of the human GAPDH gene was initiated. We gratefully acknowledge M. J. Siciliano for providing advice on GAPDH isoenzyme assays, C. Shih for providing the human genomic library, Stephanie Quamo for excellent technical assistance, Paul Habermann for sequencing the human GAPDH cDNA, J. R. Grove, R. Selden, B. Seed, and D. Moore for careful reading of the manuscript and helpful comments. L. E. remains indebted to Joseph Avruch for past and continuing support.

REFERENCES

1. Dandliker, W. B., and Fox, J. B. (1955) *J. Biol. Chem.* **214,** 275–283
2. Fox, J. B., and Dandliker, W. B. (1956) *J. Biol. Chem.* **218,** 53–57
3. Harris, I., and Perham, R. N. (1963) *Biochem. J.* **89,** 60P
4. Wang, C., and Alapovic, P. (1980) *Arch. Biochem. Biophys.* **205,** 136–145
5. Tsai, I.-H., Murthy, S. N. P., and Steck, T. L. (1982) *J. Biol. Chem.* **257,** 1438–1442
6. Alexander, M., Curtis, G., Avruch, J., and Goodman, H. (1985) *J. Biol. Chem.* **260,** 11978–11985
7. Rosen, O. (1987) *Science* **237,** 1452–1458
8. Bruns, G. A. P., Pierce, P., Regina, V. M., and Gerald, P. S. (1978) *Cytogenet. Cell Genet.* **22,** 547–551
9. Piechaczyk, M., Blanchard, J. M., Sabouty, S. R.-E., Dani, C., Marty, L., and Jeanteur, P. (1984) *Nature* **312,** 469–471
10. Tso, J. Y., Sun, X.-H., Kao, T., Reece, K. S., and Wu, R. (1985) *Nucleic Acids. Res.* **13,** 2485–2502
11. Benham, F. J., Hodgkinson, S., and Davies, K. E. (1984) *EMBO J.* **3,** 2635–2640
12. Hanauer, A., and Mandel, J. L. (1984) *EMBO J.* **3,** 2627–2633
13. Stone, E. M., Rothblum, K. N., and Schwartz, R. J. (1985) *Nature* **313,** 498–500
14. Dugaiczyk, A., Haron, J. A., Stone, E. M., Dennison, O. E., Rothblum, K. N., and Schwartz, R. J. (1983) *Biochemistry* **22,** 1605–1613
15. Michelson, A. M., Markham, A. F., and Orkin, S. H. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 472–476
16. Rigby, P. W. J., Dieckmann, M., Rhodes, C., and Berg, P. (1977) *J. Mol. Biol.* **113,** 237–251
17. Shih, C., Burke, C., Chou, M. J., Seldis, J. B., Yang, C. S., Lee, C. S., Isselbacher, K. J., Wands, J. R., and Goodman, H. M. (1987) *J. Virol.* **61,** 3491–3498
18. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517
19. Perron-Yanisch, C., Vieira, J., and Messing, J. (1985) *Gene (Amst.)* **33,** 103–119

20. Sinha, N. D., Biernat, J., McManus, J., and Koster, H. (1984) *Nucleic Acids Res.* **12,** 4539–4557
21. Chen, E. J., and Seeburg, P. H. (1985) *DNA (NY)* **4,** 165–170
22. Seed, B., and Aruffo, A. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 3365–3369
23. Devereux, J., Haeberli, P., and Smithies, O. (1984) *Nucleic Acids Res.* **12,** 387–395
24. Southern, P. J., and Berg, P. (1982) *Mol. Appl. Genet.* **1,** 327–341
25. Gorman, C., Moffat, L., and Howard, B. (1982) *Mol. Cell. Biol.* **2,** 1044–1051
26. Lopata, M. A., Cleveland, D. W., and Sollner-Webb, B. (1984) *Nucleic Acids Res.* **12,** 5707–5717
27. Selden, R. F., Howie, K. B., Rowe, M. E., Goodman, H. M. and Moore, D. (1986) *Mol. Cell. Biol.* **6,** 3173–3179
28. Wright, D. A., Siciliano, M. J., and Baptist, J. N. (1972) *Experientia (Basel)* **281,** 889
29. Feinberg, A. P., and Vogelstein, B. (1983) *Anal. Biochem.* **132,** 6–13
30. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manuel,* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
31. Mount, S. M. (1982) *Nucleic Acids Res.* **10,** 459–472
32. Dierks, P., Van Ooyen, A., Cochran, M. D., Dobkin, C., Reiser, J., and Weissman, C. (1983) *Cell* **32,** 695–706
33. Cochran, M. D., and Weissmann, C. (1984) *EMBO J.* **3,** 2453–2459
34. Fort, P., Marty, L., Piechaczyk, M., Sabrouty, S. E., Dani, C., Jeanteur, P., and Blanchard, J. M. (1985) *Nucleic Acids Res.* **13,** 1431–1442
35. Kozak, M. (1987) *Nucleic Acids Res.* **15,** 8125–8147
36. Arcari, P., Martinelli, R., and Salvatore, F. (1984) *Nucleic Acids Res.* **12,** 9179–9189
37. Allen, R. W., Trach, K. A., and Hoch, J. A. (1987) *J. Biol. Chem.* **262,** 649–653
38. Alexander, M., Lomanto, M., Nasrin, N., and Ramaika, C. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85,** 5092–5096
39. Shih, M.-C., Lazar, G., and Goodman, H. M. (1986) *Cell* **47,** 73–80
40. Stone, E. M., Rothblum, K. N., Alevy, M. C., Kuo, T. M., and Schwartz, R. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 1628–1632
41. Dani, C., Piechaczyk, M., Audiger, Y., El Sabouty, A., Cathala, G., Marty, L., Fort, P., Blanchard, J.-P., and Jeanteur, P. (1984) *Eur. J. Biochem.* **145,** 299–304
42. Blake, C. (1985) *Int. Rev. Cytol.* **93,** 149–185
43. Duester, G., Jornvall, H., and Hatfield, G. W. (1986) *Nucleic Acids Res.* **14,** 1931–1941
44. Huitorel, P., and Pantaloni, D. (1985) *Eur. J. Biochem.* **150,** 265–269
45. Caswell, A. H., and Corbett, A. M. (1985) *J. Biol. Chem.* **260,** 6892–6898
46. Kawamoto, R. M., and Caswell, A. H. (1986) *Biochemistry* **25,** 656–661
47. Rossman, M., Liljas, A., Branden, C., and Banaszak, L. (1975) in *The Enzymes* (Bayer, P. D., ed) vol. 11, pp. 61–102, Academic Press, Orlando, FL
48. Harris, J., and Waters, M. (1975) in *The Enzymes* (Boyer, P. D., ed) Vol. 13, pp. 1–49, Academic Press, Orlando, FL
49. Zakut, R., Shani, M., Givol, D., Neuman, S., Yaffe, D., and Nudel, U. (1982) *Nature* **298,** 857–859
50. deCrombrugghe, B., and Pastan, I. (1982) *Trends Biochem. Sci.* **7,** 11
51. Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R., and Dogson, J. (1980) *Cell* **20,** 555–556
52. Stuart, G. W., Searle, P. F., Chen, H. Y., Brinster, R. L., and Palmiter, R. D. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81,** 7318–7322
53. Carter, A. D., Felber, B. K., Walling, M., Jubier, M.-F., Schmidt, C. J., and Hamer, D. H. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81,** 7392–7396
54. Karin, M., Haslinger, A., Holtgreve, H., Richards, R. I., Krauter, P., Westphal, H. M., and Beato, M. (1984) *Nature* **308,** 513–519
55. Weiher, H., Konig, M., and Gruss, P. (1983) *Science* **219,** 626–631
56. Buchler, P., and Trifonov, E. N. (1986) *Nucleic Acids Res.* **14,** 10009–10026
57. Sager, R., and Kovac, P. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79,** 480–484
58. Chapman, A. B., Knight, D. M., Dieckmann, B. S., and Ringold, G. M. (1984) *J. Biol. Chem.* **259,** 15548–15555
59. Bird, A. (1986) *Nature* **321,** 209–213

Supplemental Material to:
Isolation and Complete Sequence of a Functional Human Glyceraldehyde-3-Phosphate Dehydrogenase Gene

Louis Ercolani, Brian Florence, Maria Denaro, and Maria Alexander

EXPERIMENTAL PROCEDURES

Cloning the Human GAPDH Gene: A chicken GAPDH cDNA (14) was used to screen a human liver cDNA library (15). A full length human GAPDH cDNA clone was obtained; the insert was excised and purified twice by Seaplaque agarose gel electrophoresis. The fragment was nick-translated by the method of Rigby, *et al.* (16) and used to screen 1 x 10^6 plaques from an EMBL Sau3A partial library (17). Fifteen positive clones were subjected to restriction endonuclease mapping and Southern analysis (18). A 12 kb insert from a clone with a restriction map consistent with the presence of introns, and a hybridization pattern consistent with the presence of the entire human GAPDH coding sequence was subcloned into pUC12 (pghGAPDH)(19) and used for nucleotide sequencing analysis, and expression studies as described below.

Nucleic Acid Sequencing: Restriction fragments of pghGAPDH were subcloned into either pUC12 (mp18) or M13 (mp18 or mp19) vectors. Oligonucleotide primers 15 to 17 nucleotides in length were synthesized by the phosphoramidite method on an automated Applied Biosystem 380B DNA synthesizer (20). Double stranded pUC12 templates were alkali denatured and ethanol precipitated as described by Chen and Seeburg (21). Single stranded M13 and double stranded denatured templates were heated to 100°C for 2 mins in the presence of the appropriate oligonucleotide primer. The annealing reaction was allowed to proceed for 20 mins at a temperature calculated to be 5°C below the Tm of the heteroduplexes. The heteroduplexes were extended in the presence of [35-S] alpha dATP using the Klenow fragment of DNA polymerase or reverse transcriptase as described by Seed and Aruffo (22). The sequencing reaction mixtures were denatured by heating to 100°C for 2 mins and subjected to electrophoresis on preheated 5-10% polyacrylamide gels containing 42% urea. Gels were dried and autoradiographed with Kodak XAR film at 23°C for 12-96 hrs. Overlapping sequences were assembled and analyzed using University of Wisconsin Genetics Group Software (23).

Cellular Transfection Assays: Ten µg of pghGAPDH was cotransfected with 1 µg of pSV2neo (24) into H35 hepatoma cells (a gift of John Koontz, University of Tennessee) by the calcium phosphate precipitation method of Gorman, *et al.* (25). H35 hepatoma cells containing stably integrated plasmids were selected by resistance to 300 µg /ml Geneticin (G418). Cell extracts were prepared for Western blot analysis, GAPDH enzymatic activity, and primer extension analysis as described below.

5'-flanking sequences of pghGAPDH were subcloned into pOGH, a promoterless plasmid containing a growth hormone reporter gene (a gift of Richard Selden, Massachusetts General Hospital). These constructs were transiently transfected into L cells using the DEAE dextran method (26) as modified by Selden, *et al.*(27). Growth hormone secreted into the growth media of transfected cells was measured 24-48 hrs after transfection using a radioimmunoassay kit (Nichols). The lowest reproducible level of growth hormone detected is 0.1 ng/ml. DNA uptake was normalized by co-transfection with RSV-CAT, a plasmid containing the chloramphenicol acetyl transferase reporter gene driven by the Rous Sarcoma Virus promoter (26). Results are expressed as % increase ± S.E.M. in ng/ml growth hormone secreted of constructs vs pOGH, normalized for chloramphenicol acetyl transferase activity.

Primer Extension Analysis: The cell monolayer was washed twice with phosphate-buffered saline, and total cytoplasmic RNA was extracted from these monolayers as previously described (6). Identification of the start site of transcription for rat and human GAPDH mRNA was determined by annealing 10 µg of cellular RNA with a [^{32}P]-end labelled oligonucleotide complementary to the coding region of both rat and GAPDH mRNA (5'-CCATGTAGTTGAGGTCAATG-3'). The RNA-DNA duplexes were extended in the presence of reverse transcriptase (30). The extension products were extracted with phenol, precipitated in the presence of 0.3M sodium acetate and 66% ethanol. After resuspension in distilled water the extension products were heat denatured at 100°C for 2 mins and subjected to electrophoresis on preheated 8% polyacrylamide gels containing 42% urea, dried and subjected to autoradiography at -80°C for 16 hrs. The extension products for rat GAPDH mRNA and human GAPDH mRNA were predicted to be 195 nucleotides and 205 nucleotides in length, respectively.

GAPDH Activity: The cell monolayer was washed twice with phosphate buffered saline and scraped into four volumes of 10 mM Tris-HCl, pH 7.4, 1 mM EDTA, and 1 mM 2-mercaptoethanol. The cells were homogenized by repeated passage through a 27 gauge needle. The supernatant was prepared by centrifugation at 100,000 x g for 60 mins at 4°C. The protein contents of the supernatant fractions were determined by Biorad assay. Ten µl of cell extract supernatant matched for protein was subjected to electrophoresis (10 mA/cm for 6 hrs at 0°C) on non-denaturing 1% Seakem agarose or 14% starch horizontal slab gels which contained 8.7 mM Tris-HCl, 2.9 mM sodium citrate, 0.67 mg/ml NAD at pH 7.0. The electrode buffer contained 0.13 M Tris-HCl, 0.043 M sodium citrate at pH 7.0 which was rapidly recirculated to prevent changes in pH. Following electrophoresis, GAPDH activity of cell extract supernatants was determined by histochemical staining of the slab gels as described by Wright, *et al.*(28). Photographic negatives of GAPDH tetramers enzymes separated by gel electrophoresis were scanned on an LKB densitometer.

Southern Blotting: Human placental genomic DNA, GAPDH cDNA, and cloned genomic fragments were digested with appropriate restriction endonucleases, separated by electrophoresis on a 1% agarose gel and transferred to NEN GeneScreen Plus membranes as described by the manufacturer (New England Nuclear). Membranes were pre-hybridized for 24 hrs at 42°C in the presence of 1% SDS, 1 M NaCl, 10% dextran sulfate, and 50% deionized formamide. Hybridization was performed under similar conditions in the presence of denatured probes labelled with [^{32}P] alpha ATP by priming with random hexamers followed by extension of these primers with the Klenow fragment of DNA polymerase (29). After hybridization, membranes were washed for 30 mins, twice in: 2x SSC (1x SSC = 0.15 M NaCl/0.015 M sodium citrate) at 23°C; 2x SSC/1% SDS at 65°C; then 0.1x SSC at 23°C. The membranes were dried and autoradiographed with Kodak XAR film at minus 80°C for 12-96 hrs. The number of GAPDH-related sequences in the human genome was estimated by comparison of the signal strength of genomic DNA fragments to that of multiple dilutions of cloned cDNA fragments.

RESULTS

Isolation and Sequence Analysis of Putative Human GAPDH Genomic Clones.

A human genomic library was screened with a full length human GAPDH cDNA probe. Of the 10^6 recombinant phage screened, 63 positives were detected, 15 of which were analyzed by restriction endonuclease mapping. One of these clones lacked introns and was clearly a processed pseudogene. Three independent clones had similar maps compatible with the presence of introns. Southern blot analysis of these clones with oligonucleotide probes specific for the 5' and 3'-untranslated regions of the human GAPDH cDNA suggested that one of these clones encompassed the entire coding region and one kilobase of 5'-flanking sequence in a 12 kb DNA fragment. The insert of this clone was subcloned into the BamHI site of pUC12 (pghGAPDH). Limited sequence analysis of pghGAPDH confirmed the presence of introns, a TATAA box, a translation start and stop site, and a polyadenylation signal, which suggested this clone encoded a functional gene. For further nucleotide sequence analysis and expression studies, subclones of the pghGAPDH insert were constructed based on the restriction endonuclease map seen in Fig. 1. Nucleotide sequence was obtained on both strands by sequencing overlapping subclones or sequencing the original pghGAPDH plasmid with oligonucleotide primers. Ambiguous sequences from G-C rich regions were resolved by use of reverse transcriptase.

As seen in Fig. 2 and Table 1, the isolated GAPDH gene consisted of 9 exons and 8 introns. The 5' and 3'-end of the introns conformed to the eukaryotic gene splice donor/acceptor consensus sequence AAG\| GTRAGA/ ... /Y6-11\|A (31). Typical eukaryotic signals for RNA polymerase II, TATAAA, and polyadenylation, AATAAA were present. The 5' and 3'-flanking regions of this gene were G-C rich, 65% and 56% respectively. An additional conserved promoter sequence GGCCAATCT, the "CAAT box" found in virtually all globin genes, was found 203 bases upstream of the human GAPDH TATAAA sequence (32,33). Seventy nine bases downstream of the putative transcription start site (34), a sequence similar to the Kozak consensus sequence for initiation of translation, GACACCATGG, was found (35).

In Fig. 3 the exon sequence of the human GAPDH gene (line a) was compared to published sequences for muscle (12), liver (10,36), and erythrocyte (37) human GAPDH cDNAs in addition to the human GAPDH pseudogene (12). Of note, no single base substitution or deletion in the five published GAPDH sequences was collectively confirmed by the other four sequences. The human muscle cDNA sequence (line b) differed by one base at position 468, T for C, which did not change the amino acid sequence. Examination of the two reported cDNAs for human liver GAPDH revealed a one base substitution of G for C at position 279 in the sequence of Tso, *et al.* (10), (line c) which did not alter the amino acid sequence. The published liver GAPDH cDNA sequence of Acari, *et al.* (36)(line d), showed three substitutions, A for G, G for A, and A for G at positions 746, 747, and 888 and an inversion of the 5' untranslated region from positions 21 to 53, none of which were confirmed by the other GAPDH cDNAs. Examination of the erythrocyte GAPDH cDNA sequence (line e) revealed a one base substitution of C for G at position 649 which did not change the amino acid sequence. The GAPDH pseudogene on chromosome X displayed 96% sequence identity to the human GAPDH gene exon sequence.

The Isolated Gene Contains a Functional Promoter and Encodes a Functional Polypeptide.

Restriction fragments of the human GAPDH gene containing the following 5'-flanking segments: -487/+20, -268/+20, and -112 /+20 were subcloned into pOGH, a promoterless plasmid which contains coding sequences from the human growth hormone gene (27). The construct from -268 to +20 included both TATAA and CAAT box sequences, while that from -112 to +20 was devoid of the CAAT box sequence. These constructs were transiently transfected into L cells. In three experiments, cells transfected with constructs containing the GAPDH 5'-flanking sequences extending to -112, -268, and -487 bases demonstrated a 3178% ± 180, 5038% ± 92, and 5222% ± 1142 increase in growth hormone secretion respectively, as compared to the pOGH plasmid alone which secreted 1.1 ± 0.16 ng/ml of growth hormone. These experiments indicated that the isolated GAPDH gene contained a functional promoter within -112 nucleotides of the putative transcription start site and that deletion of the CAAT box sequence had only a modest effect on gene expression.

The pghGAPDH plasmid, containing the isolated human GAPDH gene, was stably transfected into rat H35 hepatoma cells. One transfected cell line, 731 was used for the experiments described below. Total RNA was extracted from human placenta, human JEG3 choriocarcinoma cells, rat H35 hepatoma cells, and 731 cells for primer extension analysis as described in Experimental Procedures. As seen in Fig. 4, the extension product for placenta or JEG3 choriocarcinoma cells was 205 nucleotides (lanes b and c) whereas the extension product from H35 hepatoma cells was 195 nucleotides (lanes d and e). In contrast, two distinct extension products identical in length to the extension products for human and rat GAPDH mRNA were seen in 731 cells (lane a). These data indicated that transcription of human GAPDH mRNA was correctly initiated from the plasmid pghGAPDH. Densitometric analysis of an appropriately exposed autoradiogram revealed 731 cells contained 5-fold more human GAPDH mRNA than rat GAPDH mRNA.

In studies published elsewhere (38), we have demonstrated that a full length (37kd) human GAPDH polypeptide product is produced by 731 cells. To assess the functional activity of this human GAPDH polypeptide, cell extracts were subjected to electrophoresis on non-denaturing starch slab gels, Fig 5.(lanes A-C) or agarose gels (lanes D-E) at pH 7.0. Under these conditions, GAPDH enzymes from H35 hepatoma cells migrated towards the anode (lanes B and E), and purified human GAPDH enzyme from erythrocytes migrated towards the cathode (lane A). Cells transfected with pghGAPDH contained GAPDH enzymes with a complex pattern of migration consistent with the random association of both human and rodent GAPDH subunits to form enzymatically active chimeric tetramers (lane C). Chimeric GAPDH tetramers could be resolved on 1% agarose gels (lane D). These enzymes detected closest to the cathode were pure human GAPDH tetramer (HHHH), whereas chimeric tetramers of human and rodent GAPDH were detected from the cathode to the origin (HHHR, HHRR, HRRR). Pure rodent tetramers (RRRR) were not detected in 731 cells but were seen in non-transfected H35 cells (lane E) Pure erythrocyte GAPDH tetramers migrated in a pattern identical to that seen for HHHH tetramers in lane D (data not shown). Densitometric analysis of both starch and agarose gels revealed that human GAPDH polypeptides contributed 5 fold more GAPDH activity than rodent polypeptides in 731 cells, consistent with the proportionate increase in human GAPDH mRNA as compared to rodent GAPDH mRNA. These data confirmed that pghGAPDH encoded a functional human GAPDH polypeptide product.

Comparison to the Chicken GAPDH Gene.

Primordial birds and mammals diverged approximately 270 million years ago (39), nevertheless comparison of the human and chicken GAPDH gene sequences revealed striking organizational similarities (40). As seen in Fig. 6, the arrangement of exons differed only in the fusion of chicken exons VIII-XI into human exon 8. Exon lengths between the two genes were virtually identical from Exon 3 to Exon 8 (Table 1). The overall length of the two genes was similar differing only by 70 bases in length from the transcriptional start site to the polyadenylation signal. This difference was due to the creation of a larger intron B in the human sequence which is not entirely compensated by the intron losses that produced a fused exon 8. As seen in Fig. 7, dot plot comparison of the entire chicken GAPDH gene to the human GAPDH gene revealed striking sequence similarity between exons but only limited sequence similarities between introns and the 5'-flanking regions of the two genes. However, the immediate intron-exon borders did show significant similarity which may reflect a contextual conservation of those segments necessary for organization of the splicing machinery in both organisms (Table 1).

Southern Blot Analysis of the Human GAPDH Gene.

Tso, *et al.* (10)and Piechaczyk, *et al.* (9) have demonstrated that a large number of GAPDH-related sequences exist in the human genome. To determine whether the isolated GAPDH DNA segment was unique, human placental DNA was digested with either HindIII and BglI, or PstI and transferred to Gene Screen Plus membranes for Southern blot analysis. As seen in Fig. 8, a probe derived from intron B detected a 1.9 kb band in placental DNA digested with PstI (lane A) and a 1.1 Kb band in the placental DNA digested with HindIII and BglI (lane B). The sizes of these detected fragments are identical to those predicted based on the restriction map for this particular GAPDH gene. The signal intensity of in each digest indicated that one copy of this gene was present in the human genome. The same results were obtained with lymphocyte DNA isolated from a different individual (data not shown).

To ascertain whether a GAPDH gene with an intron structure more closely approximating that of the chicken GAPDH gene existed in the human genome, Southern blot analysis of human placental DNA digested with PstI was performed with a probe which contained 60 nucleotides 5' of AATAAA and the 3'-flanking region of the gene (lane G). Only one copy of the predicted 1.1 kb GAPDH gene fragment was detected in PstI digests of placental DNA; a higher molecular weight band corresponding to a copy of a gene with introns dividing Exon 8 was not found. Approximately 10 copies of an apparent 1.2 kb GAPDH pseudogene were detected with this probe. To confirm that the 1.2 kb GAPDH-related DNA segment was a processed pseudogene, a full length human GAPDH cDNA was digested with HindIII and fragments 5' or 3' to this site were used to analyze human placental DNA digested with PstI. As seen in Fig. 8 (lanes C and F), both cDNA fragments detected the 1.2 kb processed pseudogene. In contrast to the two DNA segments detected with a probe derived from the 3'-untranslated region of the gene (lane G), the pattern of bands detected with the 5' and 3'-GAPDH cDNA probes was markedly more complex (lanes C,D, E, F). These findings indicated that three distinct types of GAPDH related sequences existed in the human genome: a unique functional gene, a 10 copy processed pseudogene, and numerous other GAPDH-related genes which diverged from the other two types in their 3'-untranslated regions.

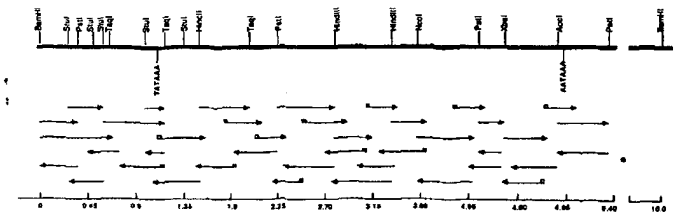

FIG. 1. *Sequencing strategy for the human GAPDH gene.*
Restriction sites that were utilized to prepare subclones for DNA sequencing are shown. Arrows indicate the direction and extent of individual sequencing runs. Arrows with boxes indicate where unique oligonucleotide primers were used to confirm restriction sites.

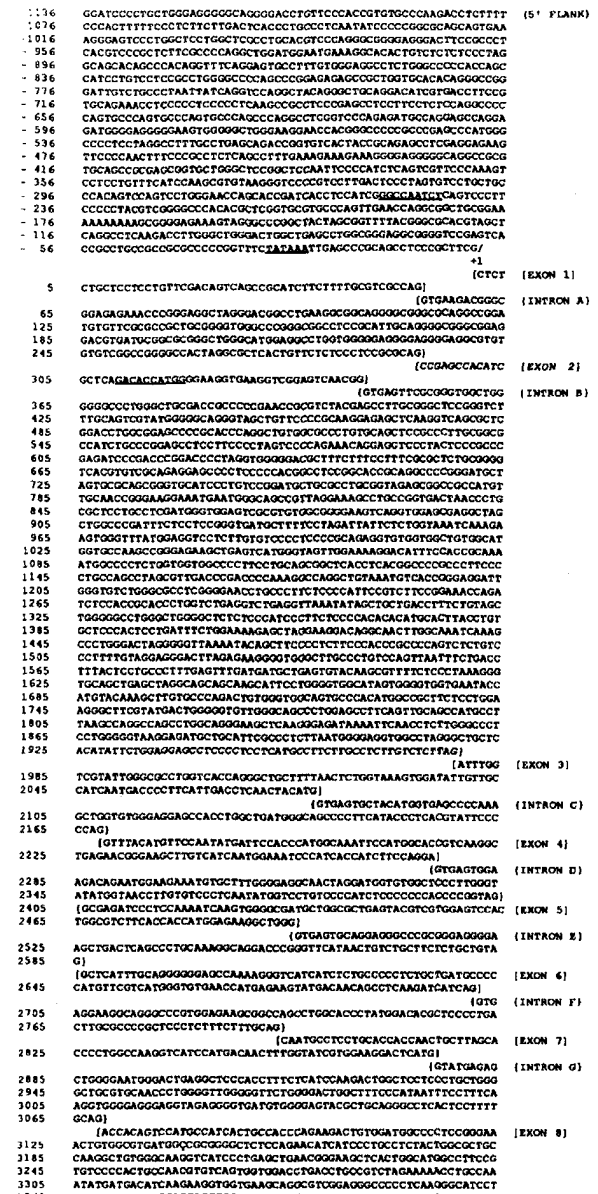

FIG. 2. *Complete sequence of the human GAPDH gene and the immediate flanking sequences.*

Underlined sequences include GGCCAATCT (CAAT box), TATAAA (RNA polymerase II entry site), GACACCATGG (Kozak consensus sequence for the translational start site),TAA (the translational stop site), and AATAAA (the polyadenylation signal). [ ] indicates exon boundaries. { } indicates intron boundaries. The complete sequence is 5377 nucleotides in length.

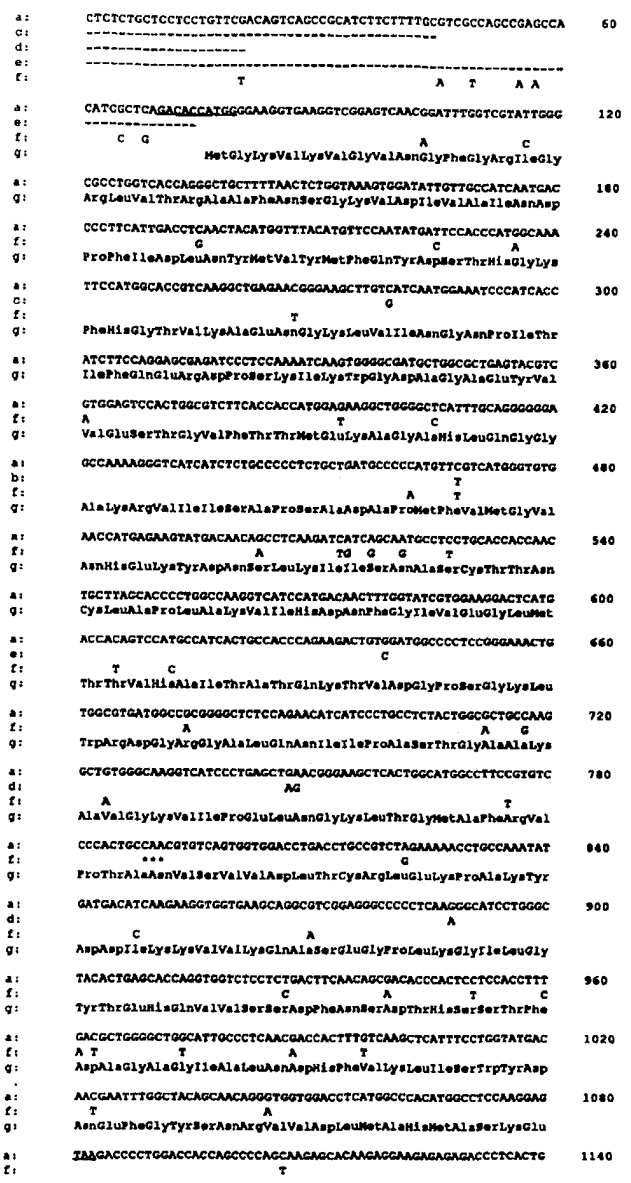

FIG. 3. *Comparison of the human GAPDH gene exon sequence to other GAPDH sequences.*

The human GAPDH gene exon sequence (a) and its deduced amino acid sequence (g), were compared to the published cDNA sequences for human muscle GAPDH (b) of Hanauer and Mandel (12), human liver GAPDH (c) of Tso, et al. (10), human liver GAPDH (d) of Acari, *et al.* (36), the human erythrocyte GAPDH (e) of Allen, *et al.* (37), and a human GAPDH pseudogene (f) of Hanauer and Mandel (12). Only base differences with the human exon sequence are shown. A bullet in the sequence (•) indicates a missing base. Dashed lines (-) indicate that a sequence was not reported in these areas. The underlined sequences include the translational start site, the translational stop site, and the polyadenylation signal. Note bases 21 to 53 of sequence d are entered as the reverse complement of the published sequence (36).

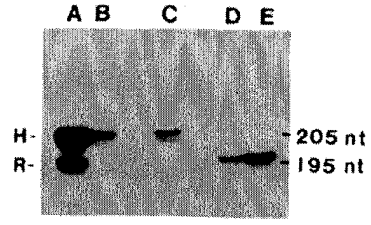

FIG. 4. ***Primer extension analysis of GAPDH mRNA.***

Total RNA was isolated from human placental tissue, human JEG3 choriocarcinoma cells, rat H35 hepatoma cells transfected with pghGAPDH (731) and untransfected H35 hepatoma cells. GAPDH mRNA was quantitated by primer extension analysis as described in Experimental Procedures. The extension product for human GAPDH mRNA designated **H** (205 nucleotides) is 10 nucleotides longer than rat GAPDH mRNA designated **R** (195 nucleotides).

Lane A: 731 cells treated with 1 mu/ml insulin for 24 hrs prior to harvesting mRNA contained two predominant extension products, 205 and 195 nucleotides in length.

Lane B: Human placenta contained a single extension product of 205 nucleotides.

Lane C: Human JEG-3 choriocarcinoma cells contained a predominant extension product of 205 nucleotides.

Lane D: H35 rat hepatoma cell contained a predominant extension product of 195 nucleotides.

Lane E: H35 rat hepatoma cells treated with 1 mu/ml insulin for 24 hrs prior to harvesting mRNA contained a predominant extension product of 195 nucleotides.

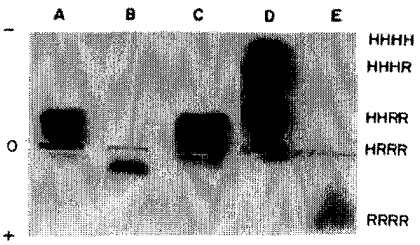

FIG 5. ***Isoenzyme analysis of GAPDH proteins.***

A 14% starch gel was used to compare the histochemical activities and migration of GAPDH tetrameric enzymes. Cell extracts were matched for protein and were subjected to electrophoresis at pH 7.0 as described in *Experimental Procedures*. (-) indicates cathode, (o) indicates origin, and (+) indicates anode.

Lane A: pure human erythrocyte GAPDH designated **Human.**

Lane B: H35 hepatoma cell extracts designated **RAT**.

Lane C: H35 hepatoma cell extracts transfected with pghGAPDH designated 731.

A 1% Seakem agarose gel was used to compare the histochemical activities and migration of GAPDH tetrameric enzymes (Lanes D and E) as described in *Experimental Procedures*. The agarose gel was aligned at its origin to that of the starch gels (Lanes A,B, and C).

Lane D: 731 cell extracts display four distinct areas of GAPDH histochemical activity corresponding to the association of human (H) and rodent (R) subunits. Pure human GAPDH activity with four human subunits **HHHH** displayed predominant cathodic migration. Chimeric enzymes with increasing rodent subunit composition **HHHR**, **HHRR**, and **HRRR** displayed decreasing cathodic migration.

Lane E: H35 hepatoma cell extracts displayed only one tetramer **RRRR** which migrated toward the anode.

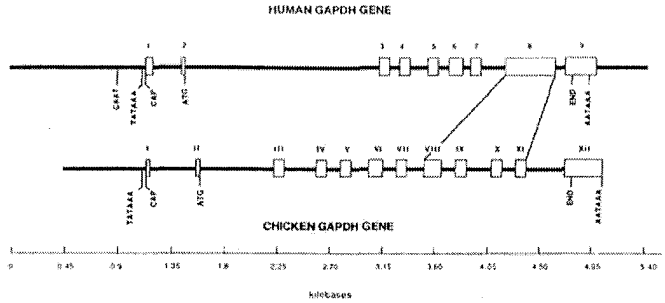

FIG 6. ***Comparison of the human and chicken GAPDH genes.***

The DNA sequence of the human GAPDH gene is shown above the chicken GAPDH gene. Open boxes indicate exons. The sequences are aligned at the putative transcription initiation site for each gene. Lines between the two genes indicate the loss of chicken introns to form a fused exon 8 in the human GAPDH gene.

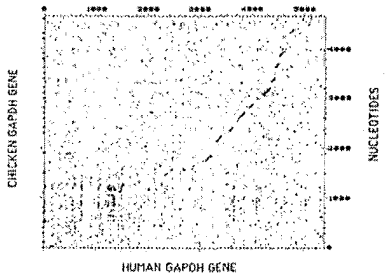

FIG 7. ***Dot Plot comparison of human and chicken GAPDH gene sequences.***

The chicken GAPDH gene (Y axis) and the human GAPDH gene sequence (X axis) were compared utilizing a Dot Plot program from the University of Wisconsin Genetics Group (23). Dots represent areas of 6 nucleotide sequence identity between the two genes.

TABLE I.

Comparison of sequences from the splice junction borders of the Human and the Chicken GAPDH Genes[1].

Exon No.	Bases	3' Exon Pur]	5' Intron (GT	-	3' Intron PyrAG)	5' Exon (X [2]-	% Sequence	Identity[3].	
I.	49	CCTCTGCCAG]	[GTGAAGACCT	-	TCCCCCCCAG]	[CCCAGCCAG	44%]	[55%	89%] [22%
I.	39	CTTCCACCAG]	[GTAACCCCCT	-	TCCCCCCCAG]	[CAACACCTAT			
II.	52	ACTCAACCCG]	[GTGAGTTTCG]	-	GTCTCTTTAG]	[ATTTGGTCGT	100%]	[89%	67%] [89%
II.	41	ACTCAACCCG]	[GTGAGTTCCG]	-	TTCTCCCCAG]	[ATTTGGCCCG			
III.	100	AACTACATCG]	[GTGAGTCGCT	-	TTCCCCCAG]	[GTTTACATG	100%]	[67%	44%] [100%
III.	100	AACTACATCG]	[GTGAGTTGC	-	CCCTGCTCAG]	[GTTTACATG			
IV.	107	CTTCCCAGGA]	[GTGAGTTCCA	-	CCCCCGTTAG]	[CCCAGATCC	100%]	[78%	33%] [78%
IV.	108	CTTCCACCCA	[GTGAGTTAGCT	-	GTTCCTCCAG]	[CCTTGACCCC			
V.	93	AAGGCTCCCC]	[GTGAGTTCCA	-	CTCCTGTTAG]	[CCTCCATTTG	100%]	[56%	44%] [89%
V.	90	AAGCCTCCCC]	[GTAAGTTACC	-	TTTTTTCAG]	[CCTCCATCTG			
VI.	116	GATCATCAG]	[GTGAGGAAG	-	TCTTTGCAG]	[CAATGCCTC	67%]	[67%	56%] [89%
VI.	116	AATTGTCAG]	[GTAAAGGAG	-	GTTTGCCAG]	[CAATGCCATC			
VII.	42	CCACTCCATG]	[GTATGAGAG	-	CTTTTGCAG]	[ACGCAGAGTC	89%]	[56%	44%] [89%
VII.	62	GCTCTTATG]	[GTAGGTCTG	-	TTTCCCTAG]	[ACCACTGTC			
VIII.-XI.	413	CATTTCCTG]	[GTATGTGCC	-	ATCTTCTAG]	[GTATGACAA	78%]	[56%	33%] [100%
VIII.-XI.	413	TGTTTCCTG]	[GTATGCACA	-	CTTCCTCAG]	[GTATGACAA			

1 The human sequence is shown above the chicken sequence from Stone, *et al.* (40) which lacked a 3' flanking sequence.
- = intervening nucleotides between the 5' and 3' ends of introns, [ ] = exon borders, { } = intron borders

2. Consensus sequence at splice junctions for the human and chicken GAPDH genes. *Pur* = purine, Pyr= pyrimidine, X= any base

3. Sequence identity between human and chicken over a nine base region are displayed. For purposes of comparison, the chicken sequence has been displaced by one base to account for a possible missing guanosine at the splice sites between exon IV - intron IV, and exon V.